

eXtremely Large Dataset Management (XLDM)

Data Foundations Interest Group

- **Focus issue:** Manage exponentially growing volumes of research data (often arriving in time-sensitive data streams)
 - Generated by networks of sensors & instruments or as output from simulations
 - Sizes from multi-terabytes to petabytes
 - Management challenges: architecture, storage, search, retrieval
 - Processing challenges: large-scale analytics, transformation, fusion

eXtremely Large Dataset Management (XLDM)

Data Foundations Interest Group

- **Approaches:**

- **Long-Term:** Build “generic” XL data management technologies or systems
- **Short-Term:** Identify diverse use cases dealing with extreme dataset sizes; Recognize best practices; Formulate benchmarks; Design query languages; Design storage & access architectures (partitioning, replication, parallelism); Build prototypes
- **Vehicle of Activities:** Int’l XLDB community of >1K data researchers, domain scientists, data practitioners & data center managers, industry; other int’l collaborations with similar constituencies

eXtremely Large Dataset Management (XLDM)

Data Foundations Interest Group

- **Alignment with RDA: WG possibilities**
 - Use cases from diverse disciplines with XL datasets to identify common requirements (possible connections with other WGs)
 - Best practices from individual communities (possible connections with other WGs)
 - Benchmarks for managing and processing XL datasets
 - Standardizing an array query language
 - Specialized policies for dealing with XL datasets (connection with policy WG)